オペレーティングシステムSUPER－UX

浜地　真 *1　岡本　明 *2　板垣　治敏 *3

1．はじめに

スーパーコンピュータはハードウェア技術の進歩によるコストパフォーマンスの向上とともに、大学や研究機関だけでなく、民間企業での新技術開発や新製品開発において重要な手段として利用されており、技術革新の変化の著しいこの時代には必須のコンピュータとなっています。

一方、スーパーコンピュータのオペレーティングシステム（ＯＳ）は、従来独自のＯＳが中心でしたが、最近はＵＮＩＸを基盤としたＯＳが主流になっています。しかしながら標準ＵＮＩＸをスーパーコンピュータに適用する場合、性能的・機能的にスーパーコンピュータにふさわしい大幅な強化が必要です。

ＳＵＰＥＲ－ＵＸはＡＴ＆Ｔ　ＵＮＩＸ　Ｓｙｓｔｅｍ　Ｖ　オペレーティングシステムに準拠し、さらに４.３ＢＳＤ機能を取り込み、その特徴を継承するとともに、互換性を保ちつつ、スーパーコンピュータにふさわしい多くの機能強化を行っています。

本稿では大幅に機能強化したマルチプロセッサ制御機能、並列処理機能、大規模・高速ファイル入出力機能、バッチ処理機能、柔軟なスケジューリング機能、運用管理機能、高信頼システム、充実したネットワーク機能などを中心に紹介します。

2．概要

ＳＵＰＥＲ－ＵＸはオープンシステムに適した業界標準のＯＳとして広く普及しているＵＮＩＸベースのＯＳです。ＳＵＰＥＲ－ＵＸはＡＴ＆Ｔ　ＵＮＩＸ　Ｓｙｓｔｅｍ　Ｖ　Ｒｅｌｅａｓｅ３.１に準拠し、大規模・超高速のスーパーコンピュータ向きに大幅に強化したＯＳです。

以下にＳＵＰＥＲ－ＵＸの特徴を述べます。

（１）　高速性の追求

１）　高速ファイル入出力機能

クラスタＩ/Ｏ、ＳＦＳ、非同期Ｉ/Ｏ、仮想ボリュームキャッシュ、並行入出力

２）　拡張記憶ファイル

３）　超高速磁気ディスク装置

（２）　大規模システム

１）　大容量ファイルＳＦＳ（２０４８Ｔバイト）

２）　最大８Ｇバイトの主記憶装置

３）　３４Ｇバイト磁気ディスク装置

４）　最大１,２５０Ｇバイトのカートリッジライブラリ

（３）　ネットワーキング

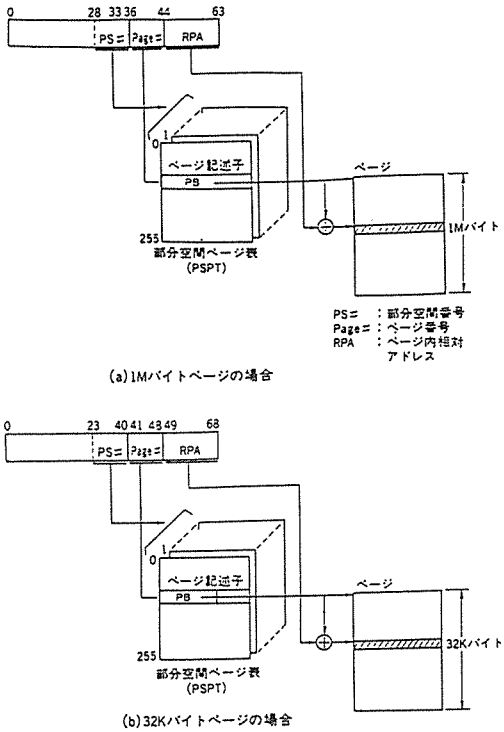

図1　アドレス変換

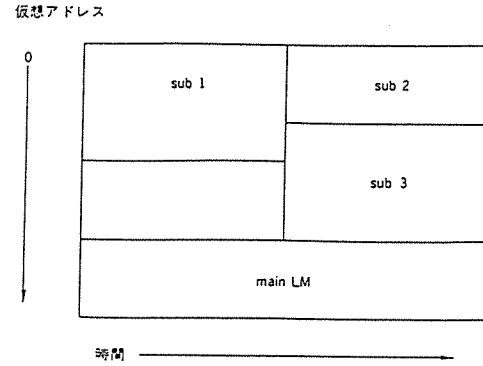

図2　LMオーバレイ

また、通常、ｕｎｉｘコマンドは３２Ｋバイトページ、高速演算用のベクトルジョブは１Ｍバイトページを使用しますので、この動的な主記憶分割はベクトルジョブの高速演算処理に外乱を与えることなく、ｕｎｉｘコマンドが使用できることを意味しています。

（３）　巨大な仮想空間

プロセス当たり１１Ｇバイトの論理空間を割り当て、最大８Ｇバイトの主記憶により、巨大プログラムの実行を可能としています。

主記憶以上のプログラムに対しては、ＬＭオーバレイ機能により実行を可能としています。

ＬＭオーバレイ機能は、ソースプログラムを修正することなく、プログラムをｍａｉｎＬＭと複数のｓｕｂＬＭに分割して仮想空間を重複させ、ｓｕｂＬＭ単位にロード／アンロードする事により、小さい主記憶でより大きなプログラムを実行可能とする機能です。

その概念図を図２に示します。

（４）　高速スワッピング

主記憶不足時のプロセス（ジョブ）の多重動作は、スワッピング機能により可能としています。

プロセスの主記憶サイズやプライオリティに応じて、主記憶を保持できる時間（メモリ滞在保障

時間）を、プロセス単位に動的に設定できますので、無駄なスワッピングを防止することができ、効率的で柔軟なシステム設計が可能となっています。

また、スワップアウトはプロセス全体を対象とするのではなく、スワップインに必要なサイズ分しか対象とせず、かつ、ページの有効サイズのみを対象としていますので、巨大ジョブの高多重度も充分に保障することができます。

さらに、XMU（拡張記憶装置）をスワップファイルにすることにより、より高速なスワッピングが可能となります。

4．柔軟なスケジューリング制御

ＳＵＰＥＲ－ＵＸのスケジューリング方式は標準ＵＮＩＸのスケジューリング方式を基本に大幅な強化を行っています。開発の基本方針は顧客の運用に応じたスケジューリングを可能とする高いチューナビリティをもったスケジューリング方式を提供することです。

元来ＵＮＩＸは会話型処理向けに開発されたＯＳです。よってそのスケジューリング方式も、利用者に対してプロセッサが公平に割り当てられるようなアルゴリズムです。

一方、ＳＵＰＥＲ－ＵＸは、ＮＱＳによるバッチ処理機能を提供しています。またスーパーコンピュータとしての高価なＣＰＵや大容量メモリなどを効率的に使用できるように、各種システム管理や運用管理を行うためのコマンドを用意しています。これらは、いわゆる一般ユーザが端末から使用するコマンドとは異なる性質をもっています。

以上のようにバッチ処理やシステム運用管理を有効に動かすことができるように、ＳＵＰＥＲ－ＵＸではメモリスケジューリングを含め、標準ＵＮＩＸのスケジューリング方式を全面的に見直しました。

（1）　ドメイン

ＳＵＰＥＲ－ＵＸでは、会話型処理ドメインとバッチ処理型ドメインを設定しています。このドメインには、

①　ＣＰＵ配分のための優先度

②　メモリ配分のための優先度

を設定することができます。利用の仕方の例を示します。

①のＣＰＵ配分では、昼間は会話型処理ドメインを優先的に、夜間はバッチ処理ドメインを優先的に、割り当てるようにすることができます。

②のメモリ配分では、一般的に小規模なメモリで短時間で処理を行う会話型処理ドメインの優先度を高く設定することにより、会話型処理ドメインがメモリ常駐されやすくするなどの設定ができます。

さらに、定期的に会話型処理とバッチ処理のＣＰＵ使用時間を監視することができる機能も提供

しています。

（2） スケジューリンググループ

各ドメインはシステムで既定のスケジューリング方式をもっています。会話型処理ドメインでは動的優先度スケジューリング方式をもち、バッチ処理ドメインでは新規に固定優先度スケジューリング方式を導入しました。

動的優先度方式はある単位時間当たりのプロセッサの使用時間をもとに優先度を計算しなおし、できるだけ各プロセスに平等にプロセッサを割り当てるように制御する方式です。基本的には標準のＵＮＩＸで提供されている方式です。

固定優先度方式はプログラムの実行中、優先度を更新しない方式です。ジョブの特性がよく解っている場合に有効な方式であり、計画的なスケジューリングが可能です。

さらに、ＳＵＰＥＲ－ＵＸは各ドメイン内で複数のスケジューリング方式をもつことを可能にしています。スケジューリングをもつ単位としてスケジューリンググループを定義しました。

たとえば会話型処理ドメイン内では、一般ユーザやシステム管理者などにわけてスケジューリンググループを設定し、各々に異なるスケジューリング方式を持たせることができます。またバッチ処理ではバッチジョブを投入するＮＱＳのキュー単位でスケジューリンググループを設定し、異なるスケジューリング方式をもたせることができます。

各ドメインとスケジューリンググループの概念図を図３に示します。

スケジューリンググループを決定するときには、優先度の基底値、優先度を計算するための式に対しての各種要因、計算契機等を設定することができます。これらのパラメータの設定により、固定優先度スケジューリング方式や、動的優先度スケジューリング方式をもたせることができます。動的優先度スケジューリング方式では優先度の計算式に対しての各種パラメータをチューニングすることにより標準ＵＮＩＸの方式に比べて、優先度の遷移を緩やかにする方式やその逆の方式をもたせることができます。

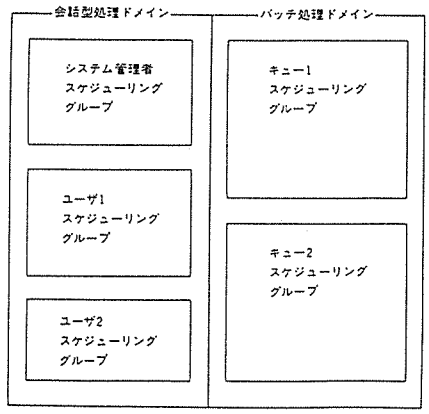

図３ ドメインとスジェジューリンググループ

以上のような機能を使用することにより、システムサイトの運用に適したスケジューリング方式の設定およびチューニングを行うことができます。

5．大規模・高速ファイルシステム

スーパーコンピュータで実行されるアプリケーションプログラムは大容量のデータを扱うプログラムが多くあります。しかしながら標準のＵＮＩＸは小規模ファイルに対する小さなサイズのファイルＩ/Ｏ向きに作られています。そこでＳＵＰＥＲ－ＵＸはファイルの大規模化・Ｉ/Ｏの高速化のために種々の機能を実現しました。そのうちの代表的な機能を紹介します。

5．1　スーパーコンピューティングファイルシステム（ＳＦＳ）

ＳＦＳ（Ｓｕｐｅｒｃｏｍｐｕｔｉｎｇ　Ｆｉｌｅ　Ｓｙｓｔｅｍ）はＳ５ＦＳ（Ｓｙｓｔｅｍ Ｖが標準装備しているファイルシステム）をベースに新規に開発した大規模・高速ファイルシステムです。

（１）　大規模ファイル（図４）

ＳＦＳにおいてもＳ５ＦＳと同様に、最小管理単位はサイズ４ＫＢの固定長の領域で分割管理されています。ＳＦＳではブロックの連続割り当ての最大単位として”クラスタ”を定義します。つまりクラスタ長（ブロック数）が割り当てる連続ブロックの最大個数となります。

また、ＳＦＳのｉノードをｓｆｉｎｏｄｅと呼びますが、その形式はＳ５ＦＳと同様です。つまり、１０個の直接アドレッシングのクラスタ指示子と、１回の間接アドレッシングのためのクラスタ指示子、３回の間接アドレッシングのためのクラスタ指示子がそれぞれ１個ずつあります。

たとえば、クラスタ長を１，０２４個（クラスタのサイズとしては４Ｍバイト）とすると次のような巨大なファイルが提供可能となります。

① 直接アドレッシング可能なファイル

サイズ＝４０Ｍバイト

② 一重間接アドレッシング可能なファイル

サイズ＝２，０４８バイト

③ 二重間接アドレッシング可能なファイル

サイズ＝2^{30}Ｇバイト

④ 三重間接アドレッシング可能なファイル

サイズ＝2^{49}Ｇバイト

このような大規模ファイルをサポートするにはファイルが複数のボリュームにまたがって構成される（マルチボリュームファイル）必要があります。これを可能にするのが仮想ボリューム機能（第２節参照）です。

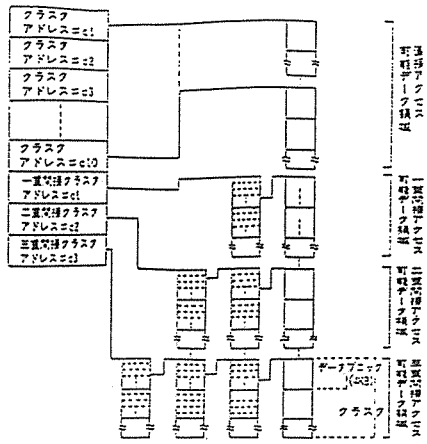

図4　大規模ファイル（ＳＦＳ）

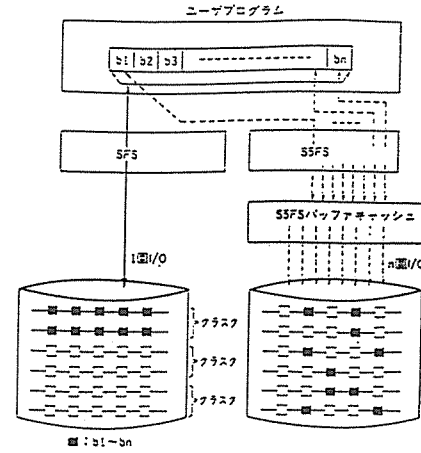

図5　クラスタＩ/Ｏ

（2）　クラスタＩ/Ｏ（図5）

Ｉ/Ｏ単位はＳ５ＦＳではブロックですが、ＳＦＳでは連続したブロックの集まりであるクラスタのサイズで一度にＩ/Ｏを行うことができます。

たとえば磁気ディスク装置で複数トラック分のサイズをクラスタとして定義した場合、主記憶装置と磁気ディスク装置の転送において磁気ディスク装置の回転ロスが生じることなく一度に転送します。不連続に存在するブロックを入出力する場合に比べ、はるかに高速な入出力が可能です。

ＳＦＳでは大容量データ転送が基本であり、Ｓ５ＦＳが提供するＵＮＩＸバッファキャッシュは利用せず、直接ユーザプログラムのデータ領域に入出力を行います。これによりＵＮＩＸバッファキャッシュが大容量のデータによってフラッシュされることなくキャッシュとして有効に働きます。またＵＮＩＸバッファキャッシュとユーザプログラムのデータ領域との大容量データ移送時のオーバヘッドもなくなり、入出力性能が高速化されます。

５．２　仮想ボリューム機能

仮想ボリューム（Ｖｉｒｔｕａｌ　Ｖｏｌｕｍｅ）は単一長ブロック（４Ｋバイト）で構成される論理的な記憶装置で、標準ＵＮＩＸのパーティションの概念、すなわちファイルシステム構築の単位を拡張したものです。

ユーザプログラムからは、あたかも次のような記憶装置があるかのようにみえます。

・大容量の記憶装置

・ストライピング記憶装置

・キャッシュ付き記憶装置

しかし実際には、これらはおのおの次の機能によって実現されています。

・マルチボリューム機能

・並行入出力機能

・キャッシュ制御機能

また、記憶装置のスペースを効率よく利用するためのリアロケーション機能を実現しています。

（1） マルチボリューム機能（図6）

標準ＵＮＩＸでは複数の磁気ディスク装置にまたがってパーティションを作成することができません。したがって作成できるファイルのサイズは物理的に１つの磁気ディスクの容量に制限されることになります。この問題を解決し複数の磁気ディスク装置からなるパーティションの作成を可能とするのがマルチボリューム機能です。これより、１台の磁気ディスク装置の容量を超えたファイルの作成が可能となります。

（2） 並行入出力機能（図7）

並行入出力機能は、主記憶装置と磁気ディスク装置とのデータ転送においてデータを分割し複数の磁気ディスク装置を同時にＩ/Ｏすることにより、Ｉ/Ｏ性能を向上させる機能です。たとえば、１Ｍバイトのデータを１台の磁気ディスク装置に書き込みするよりも、１Ｍバイトのデータを量的に４分割し、４台の磁気ディスク装置に同時に書き込みする方がより高速に処理できます。

（3） リアロケーション機能（図8）

ＳＦＳは仮想ボリュームとしての物理的な連続領域であるクラスタを領域の割り当ての単位としていますので、クラスタＩ/Ｏにより高速化が図れます。しかし逆に領域の割り当てが１クラスタに満たない場合にはスペース効率が低くなります。

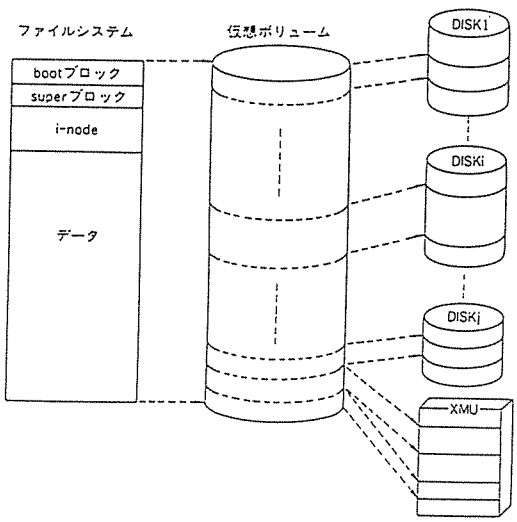

図6 マルチボリューム機能

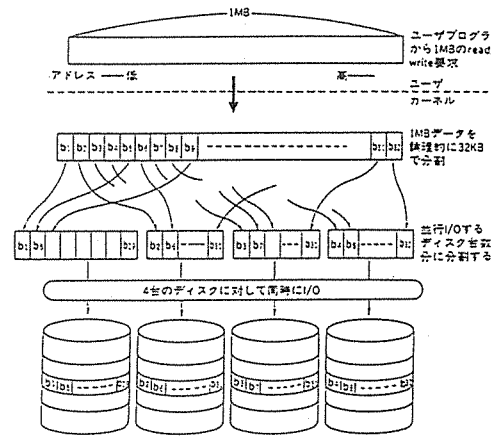

図7 並行入出力機能

そこで、クラスタを仮想化し、実際の割り当てはその単位をブロック→ステージングユニット→クラスタと移動させることによってスペースの効率化を図るリアロケーション機能を開発しました。この移動は各割り当ての単位を超えて書き込み要求が発生したときに行います。

（4）　仮想ボリュームキャッシュ機能

ある種の磁気ディスク装置には、その制御装置内にディスクキャッシュと呼ばれる高速の半導体メモリを持ち、アクセス頻度の高いデータだけをディスクキャッシュに格納して、入出力の実行性能を高めているものがあります。このディスクキャッシュ方式を磁気ディスク装置からなる仮想ボリュームに対して適用したものが仮想ボリュームキャッシュ（ＶＶキャッシュ）機能です。

仮想ボリュームは主記憶装置（MMU）、拡張記憶装置（XMU）および制御プロセッサメモリ（ＣＰＭ）から構成されます。各々AMキャッシュ、XMキャッシュおよびCMキャッシュと呼びます。仮想ボリュームキャッシュはステージングユニット（ＳＴＵ：Ｓｔａｇｉｎｇ　Ｕｎｉｔ）と呼ばれる単位で管理しています。ＳＴＵのサイズは３２Ｋバイト、６４Ｋバイトあるいは１２８Ｋバイトのいずれかの値を指定可能で、システム生成時のパラメータで指定します。各ＳＴＵはそのＳＴＵ内のデータへのアクセス頻度を示すカウンタを持っており、AMキャッシュ、XMキャッシュ、CMキャッシュおよび磁気ディスク間の移動・転送の判断基準に利用します。

次にＶＶキャッシュ機能の効果を入出力データの流れに沿って説明します。

１）　ｗｒｉｔｅ（書き込み）の場合（図９参照）

書き込み要求サイズが大きくて、大容量の一括転送（クラスタＩ/Ｏ）による高速化が期待でき

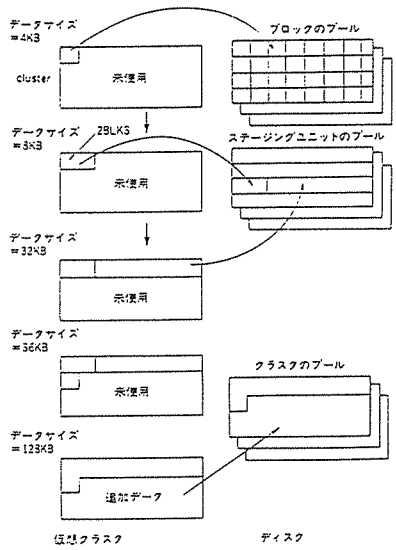

図８　リアロケーション機能

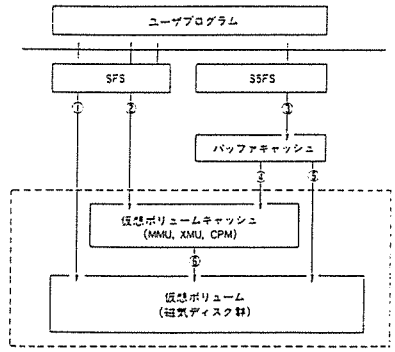

図９　書き込み

るときにはＶＶキャッシュをバイパスします（①）。要求サイズが小さい場合はＶＶキャッシュに格納します（②）。Ｓ５ＦＳの場合は必ずバッファキャッシュに格納します（③）。バッファキャッシュからＶＶキャッシュへのデータ移動（④）あるいはバッファキャッシュから仮想ボリュームへのデータ移動（⑤）が行われるのはバッファキャッシュがオーバーフローしたときと一定間隔で起動されるｓｙｎｃデーモンの実行時に行われます。このとき移動対象となるデータは、近い将来にアクセスされる可能性がほかのデータに比べて低いと判断されたデータです。次にＶＶキャッシュから磁気ディスクへの転送（⑥）はＶＶキャッシュデーモンの実行時およびＶＶキャッシュオーバーフロー時に行います。転送の対象となるデータは近い将来アクセスされないと判断される順です。

バッファキャッシュ上では４Ｋバイトのサイズにブロッキングされ、ＶＶキャッシュ上ではＳＴＵサイズにブロッキングされます。これによりＶＶキャッシュから磁気ディスク装置への転送が高速化されます。

２）　ｒｅａｄ（読み込み）の場合（図１０）

読み込み対象のデータがバッファキャッシュあるいはＶＶキャッシュ上に存在したとき、つまりキャッシュヒットしたときには、そのキャッシュから読み込みになり、高速にｒｅａｄできることになります（②、③、④）。

キャッシュ上になければ磁気ディスク装置からｓｔｕ単位で読み込みます（①、⑤）。一方、このデータはＶＶキャッシュにも格納されます（⑥、⑦）。

（５）　非同期入出力機能（図１１）

一般に入出力処理は演算処理に比べて大幅に性能が劣ります。そこで演算処理装置と入出力制御装置とが独立に実行可能であることを利用して、データの入出力処理部分と演算処理部分を並列に実行可能にして、プログラムの処理全体の実行時間を短縮することが考えられます。これを可能に

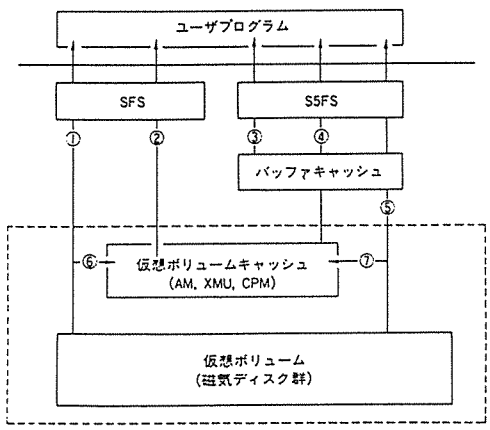

図１０　読み込み

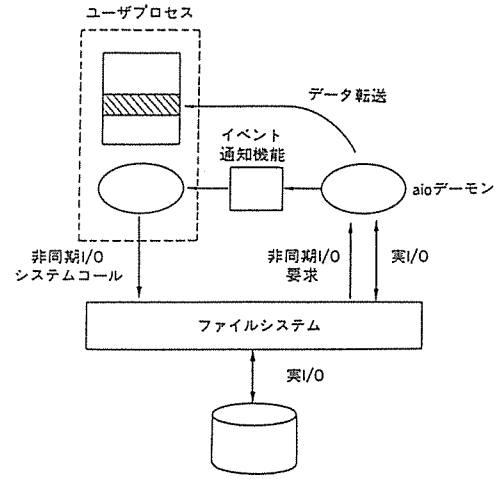

図１１　非同期入出力機能構成

するのが非同期入出力機能（Ａｓｙｎｃｒｏｎｏｕｓ　Ｉ/Ｏ）です。非同期入出力機能を使わない同期方式の場合は、Ｉ/Ｏ処理完了を待っている間ほかに演算プロセッサを使うプログラムが実行していなければ、演算プロセッサが無駄になります。非同期入出力機能を使えば、プログラムからみた見かけ上の入出力性能が大幅に向上し、また演算プロセッサを有効に使うことになりプログラムの処理時間（ＴＡＴ）が短縮されます。

６．バッチ処理

標準ＵＮＩＸは会話型処理のみを提供していますが、スーパーコンピュータのアプリケーションプログラムは実行時間が長時間かかるものが多数あり、これらはバッチ処理向きです。

ＳＵＰＥＲ－ＵＸはアメリカ航空宇宙局（ＮＡＳＡ）で開発されたバッチ処理システムであるＮＱＳ（Ｎｅｔｗｏｒｋ　Ｑｕｅｕｉｎｇ　Ｓｙｓｔｅｍ）を導入し、これをスーパーコンピュータ向けに強化しました。

ＮＱＳ（Ｎｅｔｗｏｒｋ　Ｑｕｅｕｉｎｇ　Ｓｙｓｔｅｍ）は、対話型専用ＯＳとして開発されたＵＮＩＸシステム上で、長時間ＣＰＵを使用したり、多量のメモリを使用するようなプログラムを、効率よく実行するためのバッチ処理機能を提供するシステムです。

（１）　ＮＱＳの概要

ＮＱＳはリクエストを受け付け、キューに登録し、登録したリクエストのスケジューリングを行い、順に実行していきます。

１）　リクエストとキュー

ＮＱＳのリクエストには、次の２種類があります。

①　一括して実行するプログラムを組み合わせたシェル手続きであるバッチリクエスト

②　プリンタ装置のような周辺装置に出力するデータであるデバイスリクエスト

同じようにキューには、次の３種類があります。

・バッチキュー　：バッチリクエストのみ登録可

・デバイスキュー：デバイスリクエストのみ登録可

・パイプキュー　：登録されたリクエストをほかのキューなどに転送

図１２にキュー構成の例を示します。

２）　ネットワーク機能

パイプキューの転送先は、リモートホスト上のキューでもかまいません。バッチリクエストが、リモートホストに転送され実行された場合は、その実行結果が自動的に元のローカルホストに返されます。

３）　同時実行リクエスト数の制御機能

各バッチキューやパイプキューには、同時に実行できるリクエスト数が設定できます。また、複数のキューをまとめたコンプレックスキューを定義し、コンプレックスキューとして同時に実行で

きるリクエスト数を設定することもできます。

4） 資源制御機能

バッチキューには、そのキューに登録できるバッチリクエストの、ＣＰＵやメモリなどの資源制限値を設定することができます。

この資源制限値を設定すると、バッチリクエスト登録時にバッチリクエストに指定された資源所要量と比較され、資源制限値を超えるものは登録が拒否されます。資源所要量が指定されていない場合は、資源制限値がそのバッチリクエストの資源所要量となります。

この資源所要量は、そのバッチリクエストが実行する時に、実行時の資源制限値として設定され、制限値を超えて資源を使用しようとすると実行が強制終了させられます。

5） アクセス制限機能

各キューには、そのキューにリクエストを登録することのできる利用者を制限することができます。また、利用者ごとにＮＱＳの使用を制限することもできます。これにより、予算を超過した利用者を、次の予算が付くまで、バッチリクエストを投入できなくするようなこともできます。

（2） ＮＱＳの機能

一般の利用者はＮＱＳのコマンドを使用して、以下のようなことが行えます。

① リクエストの登録、削除

② 登録済のリクエストの属性変更

③ リクエストの保留／保留解除

④ リクエストの実行中断／実行再開

⑤ リクエストの再登録

⑥ ＮＱＳのキューやリクエストの状態確認

図１３にキューの状態を表示した例を示します。

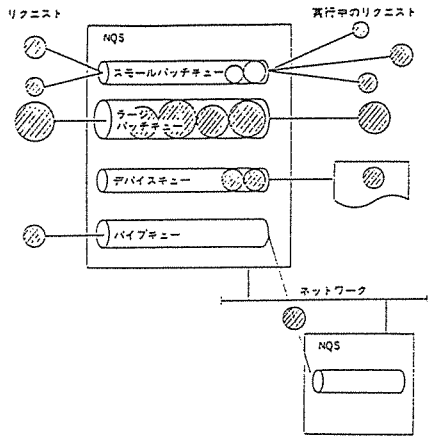

図１２ キュー構成の例

```
%qstatq

----------------------------------------------
NQS (R01.00) BATCH  QUEUE SUMMARY  HOST: sx3
----------------------------------------------

QUEUE NAME      ENA STS PRI/BPR/ TMS /MPR RLM   TOT QUE RUN WAI HLD SUS ARR
------------ --- --- ------------------ --- ---------------------------
bat_sml         ENA RUN  30/ 45/  500/ 10  3    5   2   3   0   0   0   0
bat_lrg         ENA RUN  20/ 50/ 1000/ 20  1    4   1   1   1   1   0   0
------------ --- --- ------------------ --- ---------------------------
   <TOTAL>                                      9   3   4   1   1   0   0
------------ --- --- ------------------ --- ---------------------------

----------------------------------------------
NQS (R01.00) PIPE   QUEUE SUMMARY  HOST: sx3
----------------------------------------------

QUEUE NAME      ENA  STS  PRI  RLM   TOT QUE RUN WAI HLD ARR
------------ --- ---- ---- --- ---------------------------
pipe1           ENA  RUN  30   2     1   0   1   0   0   0
------------ --- ---- ---- --- ---------------------------
   <TOTAL>                    2     1   0   1   0   0   0
------------ --- ---- ---- --- ---------------------------

----------------------------------------------
NQS (R01.00) DEVICE  QUEUE SUMMARY  HOST: sx3
----------------------------------------------

QUEUE NAME      ENA  STS  PRI   TOT QUE RUN WAI HLD ARR
------------ --- ---- ---- ---------------------------
dev1            ENA  RUN   30   3   2   1   0   0   0
------------ --- ---- ---- ---------------------------
   <TOTAL>                    3   2   1   0   0   0
------------ --- ---- ---- ---------------------------
```

図１３ キューの状態表示例

また、ＮＱＳの管理者は、ＮＱＳの管理用コマンドを使用して、以下のようなことが行えます。

① ＮＱＳ環境パラメータの設定、変更、参照

② ＮＱＳ管理者の設定、変更、参照

③ 各キューの作成、削除、参照、属性変更

④ ＮＱＳネットワーク環境の定義、変更、参照

⑤ ＮＱＳの起動／停止

⑥ 各キューの起動／停止

⑦ 各リクエストの属性変更、削除、移動

７．運用管理機能

標準ＵＮＩＸを大規模なスーパーコンピュータに適用するためには運用管理機能の充実が必要となります。

ここでは運用管理を実現している特徴的ないくつかの機能を説明します。

７．１　システムフリーズ・リスタート

運用管理者が任意の時点で、コマンドにより運用中のすべてのプロセスの実行を中断し、使用中の主記憶と拡張記憶装置（ＸＭＵ）の内容をフリーズファイルに格納し、後刻、中断時点にフリーズファイルから主記憶と拡張記憶装置を復元することにより、システムの再開を可能とするシステムフリーズ・リスタート機能を開発しました。これにより、実行時間が長時間かかるプログラムが複数個動作しているシステム環境下で、ハードウェアの定期保守や技術者の生産活動に支障を与えない運用が可能となります。

７．２　自動運転システム

自動運転システムは、ＳＵＰＥＲ－ＵＸの持つ大規模なコンピュータシステムを効率よく運用するために、運用の自動化、省力化を目的として開発しており、ＡＯＣ装置（Ａｕｔｏｍａｔｉｃ Ｏｐｅｒａｔｉｏｎ Ｃｏｎｔｒｏｌｌｅｒ）を中心としたハードウェア装置群とＳＵＰＥＲ－ＵＸの持つ諸機能とを連携して、種々の機能を実現しています。

７．３　ＯＷＳ（Ｏｐｅｒａｔｏｒ Ｗｏｒｋｓｔａｔｉｏｎ Ｓｏｆｔｗａｒｅ）

ＳＵＰＥＲ－ＵＸでは、センター運用を考慮してシステム管理者がシステム状態を集中管理しやすいようにＯＷＳを開発しました。ＯＷＳはマンマシンインタフェースに優れた「ポイント・アンド・クリック」方式のグラフィックユーザインタフェースを採用しています。

7．4　MTS（Magnetic Tape I/O Subsystem）

標準ＵＮＩＸにおいては磁気テープ装置および磁気テープ媒体の管理はユーザ自身で行わなければなりません。つまり磁気テープの排他的利用は人手に任されており、データ保護の観点から非常に危険です。

ＳＵＰＥＲ－ＵＸは磁気テープ（ＣＧＭＴ、ライブラリ型ＣＧＭＴを含む）の装置、および媒体に対する大幅な機能強化を行いました。

（1）　装置の予約・解放機能

コマンドおよびライブラリにより、使用する装置の台数、装置特性（記録密度、リワインドの有無）を指定し、装置の予約・解放を可能としています。

（2）　媒体の装置へのマウント・アンマウント機能

コマンドおよびライブラリにより、予約した装置に媒体をマウント・アンマウントする機能を有しています。

（3）　媒体へのアクセス権の妥当性チェック機能

他人（別のユーザＩＤを持つもの）がマウントした媒体には書き込み／読み込みを行うことができません。

（4）　マルチボリュームファイル

複数の媒体にまたがるファイルへのアクセスを可能としています。

（5）　ラベル処理と自動ボリューム認識（ＡＶＲ）機能

磁気テープ媒体を装置にマウントするとＡＶＲ機構が働き、ラベルの有無を確認します。ラベルがＡＮＳＩ形式かＩＢＭ形式であれば、媒体へのアクセスはそのラベル形式に従って処理します。

7．5　高速信頼システム

標準のＵＮＩＸは個人あるいは小人数で使用するパソコンやミニコン用に開発されたことから、高信頼・高稼働性が十分とは言えません。スーパーコンピュータでは、多数の人が利用する計算センター運用や２４時間連続運転運用が一般的であり、高信頼・高稼働性が要求されます。

8．高速ネットワーク

ＳＸ－３Ｒシリーズのようなスーパーコンピュータ上で実行される超大型計算では一般に大量の入力データを外部からＳＸシステムに投入したり、結果を高速に外部へ出力するためには、高速のネットワークグラフィックとして出力するサイエンティフィック・ビジュアリゼーション（Ｓｃｉｅｎｔｉｆｉｃ　Ｖｉｓｕａｌｉｚａｔｉｏｎ）の概念は、非常に重要であると考えられています。フルカラーのアニメーションを出力できるだけの帯域幅を持つ超高速ネットワークはこれを実現するための強力な手段となります。

また、計算センターの中央マシンとしてのスーパーコンピュータは、サイト内だけでなく遠方の不特定のユーザにも高度なサービスを提供する必要があります。この時、通信相手の機種にかかわらず自由に相互接続が可能であること、ネットワークのトポロジーの変化に容易に追随できることが条件となります。

大規模な計算センターや研究所などではすでにＬＡＮが張り巡らされ、ネットワーク上に各種サーバが稼働しているところも多くなってきています。ＳＸ－３Ｒシリーズはこのようなマルチベンダ環境の中に異和感なく溶け込み、超高速コンピュートサーバとしての役割を果たすことが期待されています。

８．１　開発方針

ＳＵＰＥＲ－ＵＸのネットワーク機能は、以上のような要求に応えるために次のような方針で開発しています。

（１）　高速ＬＡＮによる高速大容量通信の実現

① ＵｌｔｒａＮｅｔ

② ＦＤＤＩ（ＬＯＯＰ６７８０）

③ Ｅｔｈｅｒｎｅｔ（ＢＲＡＮＣＨ４６８０Ⅱ）

（２）　標準プロコトル／標準ＡＰＩ（Ａｐｐｌｉｃａｔｉｏｎ　Ｐｒｏｇｒａｍ）環境への対応

① ＴＣＰ/ＩＰ（プロコトルファミリ）

② ソケットインタフェース

③ ＴＬＩ（Ｔｒａｎｓｐｏｒｔ　Ｌｅｖｅｌ　Ｉｎｔｅｒｆａｃｅ）

（３）　運用管理の標準化・自動化の推進

① インタネット標準の積極的取り組み

② Ｓｎｍｐｄ，ｇａｔｅｄなど、最新のネットワーク管理技術のサポート

８．２　ＬＡＮのサポート

（１）　Ｅｔｈｅｒｎｅｔ（Ｃ&Ｃ－ＮＥＴ　ＢＲＡＮＣＨ４６８０Ⅱ）

パーソナルコンピュータを始めとして、ワークステーション、メインフレームなどほとんどすべての機器でサポートされていますので、最も手軽にネットワークを構築することができます。ＳＵＰＥＲ－ＵＸではＴＣＰ/ＩＰプロコトルをサポートしています（図１４）。

（２）　ＦＤＤＩ（Ｃ&Ｃ－ＮＥＴ　ＬＯＯＰ６７８０）

ＦＤＤＩは１００Ｍｂｐｓのトークンリング型ＬＡＮで、マルチベンダ環境でのバックボーンＬＡＮとして主流となっています。ＳＸ－３ＲシリーズとＦＤＤＩとはＬＡＮＰ（ＬＡＮプロセッサ

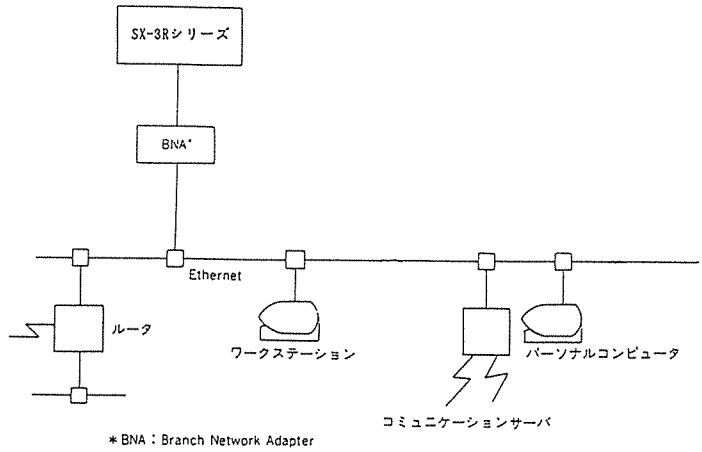

図１４　Ｅｔｈｅｒｎｅｔを利用したシステム構成例

）によって接続されます。ＬＡＮＰは、ＦＤＤＩリング上のＯＣＵ（Ｏｐｔｉｃａｌ　Ｃｏｎｃｅｎｔｒａｔｏｒ　Ｕｎｉｔ）にシングルアタッチメントステーションとして加入します。ＳＵＰＥＲ－ＵＸはＦＤＤＩ上でＴＣＰ/ＩＰプロコトルをサポートしており、ＦＤＤＩ上の他システムと高速に大量のデータを送受信することができます（図１５）。

（３）　ＵｌｔｒａＮｅｔ

ＵｌｔｒａＮｅｔはＵｌｔｒａＮｅｔｗｏｒｋ　Ｔｅｃｈｎｏｌｏｇｉｅｓ社の超高速ＬＡＮで最大１Ｇｂｐｓの速度をもっています。ＵｌｔｒａＮｅｔは、Ｈｕｂと呼ばれる交換装置をケーブルでメッシュ状に相互接続したスター型のＬＡＮの一種です。ＳＸ－３Ｒシリーズは、Ｕｌｔｒａ１０００という大型のＨｕｂにＨＩＰＰＩ（Ｈｉｇｈ　Ｐｅｒｆｏｒｍａｎｃｅ　Ｐａｒａｌｌｅｌ　Ｉｎｔｅｒｆａｃｅ）チャネルで接続されます。これにより、ＵｌｔｒａＮｅｔに接続されたグラフィックワークステーションやほかのスーパーコンピュータと画像などの大量データを瞬時にやりとりすることができます。

特に、Ｈｕｂに直結されるフレームバッファを用いれば、フルカラーのアニメーションを表示することも可能となり、サイエンティフィック・ビジュアリゼーションの実現に大いに威力を発揮します（図１６）。

８．３　ＷＡＮのサポート

（１）　ｔｔｙ回線

ＳＵＰＥＲ－ＵＸは、標準のｔｔｙドライバを備えており、最高１９．２Ｋｂｐｓまでの全二重・調歩無手順回線をサポートします。

（２）　他の広域網

専用線、高速ディジタル回線、回線交換網、Ｘ．２５パケット交換網、ＩＳＤＮ，フレームリレ

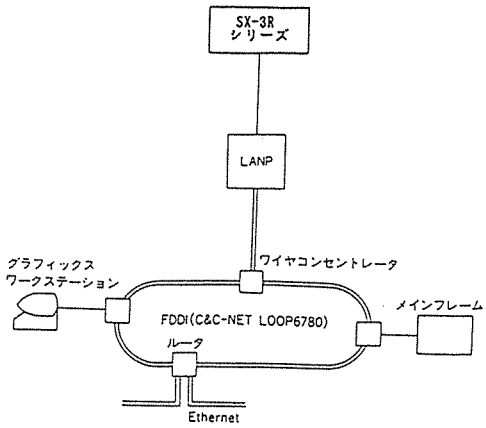

図１５　ＦＤＤＩを利用した
ネットワーク構成例

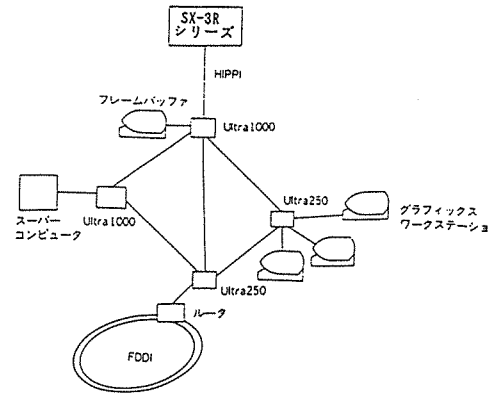

図１６　ＵｌｔｒａＮｅＴを利用した
ネットワーク構成例

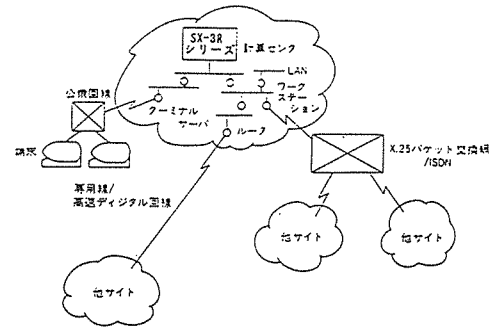

図１７　ＷＡＮを利用したネットワーク構成例

ーなどとＳＸ－３Ｒシリーズとは、ルータあるいはワークステーションを介して接続されます。この場合。各種の広域網は、分散したサイトのＬＡＮを相互に結合するという役割を果たします。ＳＵＰＥＲ－ＵＸでは、遠方のユーザがこれらの広域網を通して使用できるように各種ルータと密接に協力するため、各種ルーティングプロコトルを備えています（図１７）。

８．４　インタネットプロトコル

ＴＣＰ/ＩＰプロトコルを含むインタネットプロトコルは、ＳＵＰＥＲ－ＵＸのメインプロトコルとして位置づけられ、フルサポートが行われています。また、変化の激しいインタネットの技術ですが、ＳＵＰＥＲ－ＵＸは常に最新のプロトコル規格（ＲＦＣ：Ｒｅｑｕｅｓｔ　Ｆｏｒ　Ｃｏｍｍｅｎｔｓ）に追随します。これにより、ＳＸ－３Ｒシリーズは本格的インタネットワーク環境の中でも十分に力を発揮し続けることが可能となっています。

ＳＵＰＥＲ－ＵＸのインタネットプロトコルの実装は、４．３ＢＳＤ－ｔａｈｏｅ版を基にしています。

更にＳＵＰＥＲ－ＵＸでは、スーパーコンピュータにふさわしい機能と性能のために、

（１）　ＩＰの理論限界である６４Ｋバイトのデータグラムを ＴＣＰ，ＶＤＰ，ＩＰプロトコルで扱えるようにする

（２）　最大６４Ｋバイトのバッファを使用可能とする

（３）　スーパーコンピュータをルータとして使用できないようにＩＰデータグラムの中継機能を抑制する

（４）　強制的に大きなＭＳＳ（Ｍａｘｉｍｕｍ　Ｓｅｇｍｅｎｔ　Ｓｉｚｅ：最大セグメントサイズ）でＴＣＰの通信を行う

などのオプションが利用可能となっています。

８．５　ＡＰＩ

ネットワークアプリケーションを構築するためのＡＰＩ（Ａｐｐｌｉｃａｔｉｏｎ　Ｐｒｏｇｒａｍｍｉｎｇ　Ｉｎｔｅｒｆａｃｅ：アプリケーションプログラミングインタフェース）として、ＳＵＰＥＲ－ＵＸは次のものをサポートしています。

（１）　ソケット

４．３ＢＳＤのソケットシステムコールを実装しています。また、ｓｅｌｅｃｔシステムコール、ｐｔｙドライバ、ｔｔｙに対する豊富なｉｏｃｔｌ機能を備えていますので、４．３ＢＳＤのネットワークアプリケーションは極めて容易にＳＵＰＥＲ－ＵＸ上で動作させることができます。

（２）　ＴＬＩ（Ｔｒａｎｓｐｏｒｔ　Ｌｅｖｅｌ　Ｉｎｔｅｒｆａｃｅ）

ＵＮＩＸ　Ｓｙｓｔｅｍ　Ｖの標準機能として、ＴＬＩライブラリをサポートしており、ＴＣＰ/ＵＤＰプロトコルを利用可能です。

（３）　ＲＰＣ/ＸＤＲ（Ｒｅｍｏｔｅ　Ｐｒｏｃｅｄｕｒｅ　Ｃａｌｌ/ｅＸｔｅｒｎａｌ Ｄａｔａ　Ｒｅｐｒｅｓｅｎｔａｔｉｏｎ）

Ｓｕｎ　Ｍｉｃｒｏｓｙｓｔｅｍｓ社で開発されたＮＦＳ（Ｎｅｔｗｏｒｋ　Ｆｉｌｅ　Ｓｙｓｔｅｍ）やＮＩＳ（Ｎｅｔｗｏｒｋ　Ｉｎｆｏｒｍａｔｉｏｎ　Ｓｅｒｖｉｃｅ）の基盤をなすＲＰＣ/ＸＤＲの機能をライブラリとして提供しています。

（４）　Ｘｌｉｂ/Ｘｔ（Ｘ　ｌｉｂｒａｒｙ/Ｘ　ｔｏｏｌｋｉｔ）

Ｘ　Ｗｉｎｄｏｗシステムがサポートされており、Ｘクライアントの移植／作成が容易に行えます。

（５）　ＯＳＦ/Ｍｏｔｉｆ

ＯＳＦ/Ｍｏｔｉｆのウィジェットやｕｉｌ（Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ　Ｌａｎｇｕａｇ

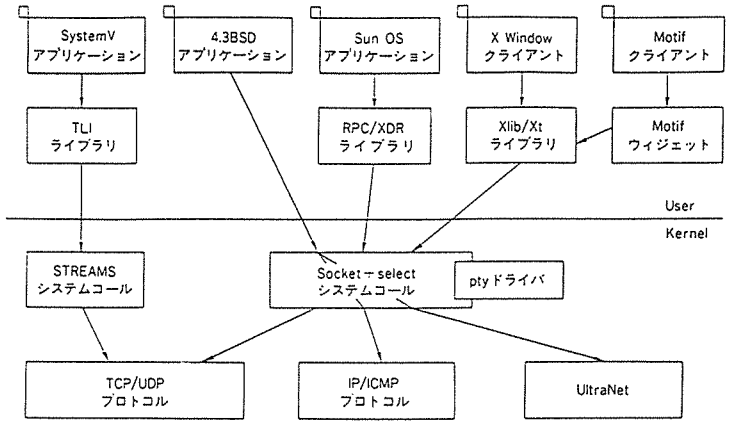

図１８　ＳＵＰＥＲ－ＵＸ　ＡＰＩサポート

ｅ：ユーザインタフェース言語）を備えておりＭｏｔｉｆクライアントの移植／作成が容易に行えます（図１８）。

８．６　分散処理

ＳＸ－３Ｒシリーズ/ＳＵＰＥＲ－ＵＸシステムは、スタンドアロンで使用しても十分な機能を備えています。しかし、計算センターの環境の中でＳＸ－３Ｒシリーズの超高速演算性能を最大限に発揮するために、ＳＵＰＥＲ－ＵＸはサーバ＝クライアントモデルに基づいた水平分散システムの概念を強力にサポートしています。計算サーバとしてのＳＸ－３Ｒシリーズは、ＬＡＮやＷＡＮで相互に接続された各種サーバやワークステーションと一体となって計算システムを構成します。ＳＵＰＥＲ－ＵＸでは、このために様々なサーバやクライアントをサポートしています（図１９）。

（１）　ｔｅｌｎｅｔ/ｒｌｏｇｉｎ

ワークステーションからＳＸ－３Ｒシリーズにログインし、対話的な利用を可能とします。

（２）　ｆｔｐ/ｒｃｐ/ｒｄｉｓｔ

ワークステーションとＳＸ－３Ｒシリーズ間で任意のファイルを転送します。ｒｄｉｓｔはファイルの配布を行います。

（３）　ＮＦＳ

ワークステーションとＳＸ－３Ｒシリーズ間で互いに相手のファイルをあたかも自分のファイルのように使用できます。

（４）　ＮＩＳ

ユーザ名やネットワーク定義の一括管理を行います。

（５）　ＢＩＮＤ（ｎａｍｅｄ/ｒｅｓｏｌｖｅｒ）

ネットワーク上のホストアドレスやメールアドレスをダイナミックに問い合わせるディレクトリ

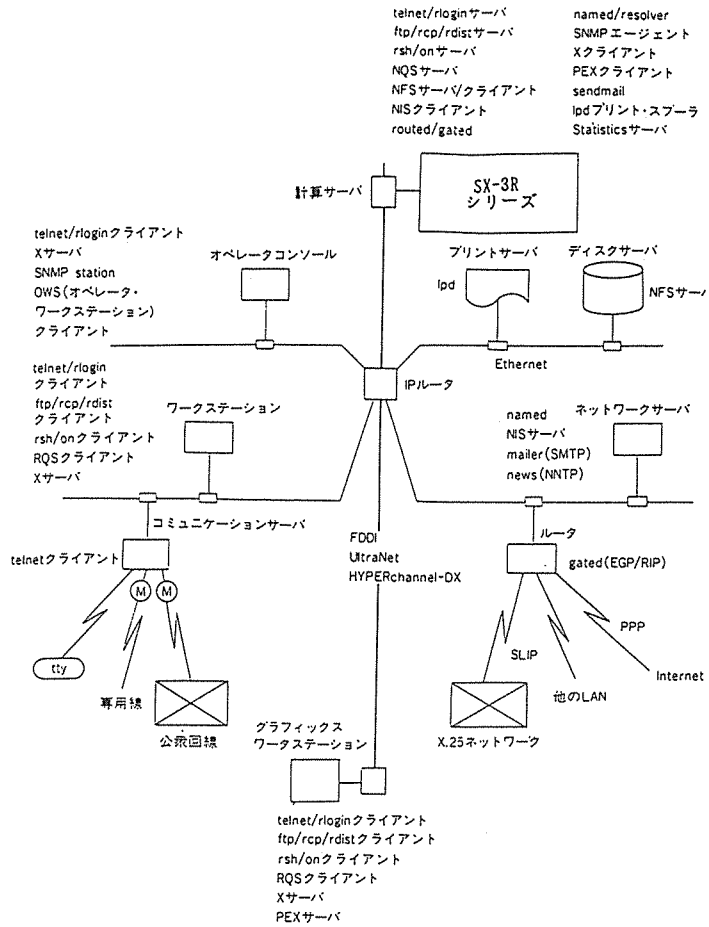

図19　SUPER－UXによる分散処理

システムの一種です。

（6）　X　Windowシステム

ワークステーションやXターミナルのビットマップディスプレイにSX－3Rシリーズ上のアプリケーション（クライアントと呼ぶ）がウィンドウを描画できます。

（7）　OWS（Operator　Work　Station）

ワークステーションがSX－3Rシリーズの専用コンソールとしてウィンドウ環境下で動作します。

（8）　rsh/on

ワークステーションとSX－3Rシリーズで互いに相手のマシン上で任意のコマンド実行する機能です。

（9）　gated/routed

ＥＧＰ，ＲＩＰ，ＨＥＬＬＯなどのプロトコルを使用してネットワークのルートテーブルを自動的に設定／更新し、オペレータや管理者の負担を軽減します。

(10)　ＮＱＳ

ワークステーションやメインフレームからＳＸ－３Ｒシリーズ上のＮＱＳにバッチＪＯＢを投入し、結果を回収することができます。また、ＪＯＢの監視・制御もネットワーク上から行えます。

(11)　ｍａｉｌ

標準のＳＭＴＰプロトコルを用いて、ネットワーク上の任意の相手と電子メールをやりとりすることができます。

(12)　ＳＮＭＰ

ＳＸ－３Ｒシリーズを含めてネットワーク上の各種ホスト／ルータ／ブリッジなどがＳＮＭＰステーションによって統合的に一括管理できます。

(13)　ｌｐｄ

ｌｐｄはネットワーク対応のスプーリングシステムであり、他のワークステーションに接続されたプリンタや専用のネットワークプリンタにデータを出力することができます。

［執筆者紹介］　＊１　NEC　第二基本ソフトウェア開発本部第三開発部
＊２　NEC　第二基本ソフトウェア開発本部第三開発部
＊３　NEC　スーパーコンピュータ販売推進本部